MANAGEMENT SCIENCE
Vol. 54, No. 6, June 2008, pp. 1115–1131
ISSN 0025-1909 | EISSN 1526-5501 | 08 | 5406 | 1115

DOI 10.1287/mnsc.1070.0832

Strategic Capacity Rationing to Induce Early Purchases

Qian Liu
Department of Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, qianliu@ust.hk

Garrett J. van Ryzin
Graduate School of Business, Columbia University, New York, New York 10027, gjv1@columbia.edu

Dynamic pricing offers the potential to increase revenues. At the same time, it creates an incentive for customers to strategize over the timing of their purchases. A firm should ideally account for this behavior when making its pricing and stocking decisions. In particular, we investigate whether it is optimal for a firm to create rationing risk by deliberately understocking products. Then, the resulting threat of shortages creates an incentive for customers to purchase early at higher prices. But when does such a strategy make sense? If it is profitable to create shortages, what is the optimal amount of rationing risk to create? We develop a stylized model to study this problem. In our model, customers have heterogeneous valuations for the firm's product and face declining prices over two periods. Customers are assumed to have identical risk preferences and know the price path and fill rate in each period. Via its capacity choice, the firm is able to control the fill rate and, hence, the rationing risk faced by customers. Customers behave strategically and weigh the payoff of immediate purchases against the expected payoff of delaying their purchases. We analyze the capacity choice that maximizes the firm's profits. First, we consider a monopoly market and characterize conditions under which rationing is optimal. We examine how the optimal amount of rationing is affected by the magnitude of price changes over time and the degree of risk aversion among customers. We then analyze an oligopoly version of the model and show that competition reduces the firms' ability to profit from rationing. Indeed, there exists a critical number of firms beyond which a rationing equilibrium cannot be supported.

Key words: dynamic pricing; capacity rationing; customer behavior; revenue management
History: Accepted by Gérard Cachon, operations and supply chain management; received October 18, 2005. This paper was with the authors 1 year and 4 months for 3 revisions. Published online in *Articles in Advance* April 25, 2008.

1. Introduction

Varying prices over time is a natural way for firms to increase revenue in response to uncertain and fluctuating market conditions. Apparel retailing is a canonical example. Demand for apparel is affected by factors such as weather, fashion trends, and economic conditions, all of which are highly uncertain; hence, forecasting demand is inherently difficult. Moreover, lead times for design, production, and distribution are normally longer than the selling season, so retailers must commit to their order quantities in advance of observing sales. As a result, they often end up with some popular products which sell out fast, whereas other unpopular products languish. In response, retailers dynamically change prices—maintaining full prices for their best-selling items while marking down slow sellers over time.

Although most retailers still make such pricing decisions manually, many are now deploying sophisticated modeling and optimization software to help support pricing decisions. (See Talluri and van Ryzin 2004, Chap. 5.) Such systems have proved quite effective; Ann Taylor, a U.S. women's apparel retailer with over 580 stores nationwide, reported a year-on-year increase in sales by 26% over the Christmas period of 2003 after it implemented markdown optimization software. Success stories like this have led to increased acceptance of model-based approaches to pricing among major retailers.

At its core, any dynamic pricing system is based on a model of demand; that is, a model of how demand responds to price changes. A typical one is to formulate demand at each point in time as only a function of the price charged at that time, the implicit assumption being that customers do not anticipate future prices; that is, they are *myopic* and buy if the current price is less than their reservation price. This model is pervasive; most commercial pricing software uses it, and it is common in the research literature too. For example, see Gallego and van Ryzin (1994), Feng and Gallego (1995), Federgruen and Heching (1999), and Chen and Simchi-Levi (2003). The myopic customer assumption is reasonable when customers make impulse purchases and for consumable goods (food, beverages, etc.). In addition, it has the considerable practical advantage of leading to mathematically tractable models.

Yet it is arguably more realistic to model customers as behaving strategically. Faced with dynamic prices, customers may benefit from strategizing over the timing of their purchases. This is especially true in the case of *durable goods*—goods that customers buy once (or more realistically, infrequently) and use over an extended period of time. When purchasing durable goods, customers are more patient and may accelerate or postpone purchasing to obtain a lower price. In his classic work, Coase (1972) gives real estate as an example of a durable good; Bulow (1982) gives the example of diamonds. Even without such extreme cases of durability, goods with a finite life can also be categorized as "durable" provided they are purchased only infrequently. Many important retail categories—apparel, sporting goods, customer electronics and appliances—can be considered durable goods in this sense. Bulow (1982) calls these *intermediate durable goods*.

Retailers certainly recognize the fact that customers often act strategically when purchasing such goods. Executives at Federated Department Stores, America's largest operator of department stores, have lamented to us that the department store industry's habit of running frequent promotional sales has "trained our customers to only buy on sale." Such behavior is evident in aggregate data too; for example, at Wal-Mart, the world's largest retailer, holiday season sales—including after-Thanksgiving and post-Christmas season sales events—account for close to 20% of total annual sales (Rozhon 2005).

Some retailers have adopted strategies specifically aimed at thwarting such strategic behavior. For example, Zara, one of the largest Spanish apparel retailers, is known for deliberately setting low stock levels for its products to encourage customers to buy when they first see products they like, rather than waiting for sales (Ferdows et al. 2005). Industry analysts estimate that unsold products at Zara represent less than 10% of stock, compared with the industry average of almost 20% as a result of this strategy (Ferdows et al. 2005).

Academic researchers too have investigated the mistakes that can occur if firms incorrectly model strategic customers as myopic. Aviv and Pazgal (2008) report extensive numerical examples showing revenue losses of up to 20% by ignoring strategic behavior. Besanko and Winston (1990) assess that the lost profit from ignoring customer strategic behavior could skyrocket to 60%.

1.1. Market and Behaviorial Assumptions

For all of these reasons, understanding strategic customer behavior and its impact on pricing and quantity decisions is important. Yet there are a wide variety of assumptions one can make in modeling customer strategic behavior, each of which has important theoretical and practical implications. Here we review these assumptions in detail and relate them to the body of research on dynamic pricing in the presence of strategic customers. We also position our work relative to these assumptions and literature.

1.1.1. Commitment to Prices. One important issue is whether a seller commits to its price schedule or whether a subgame perfect equilibrium (SPE) is a more plausible assumption. The earliest work on this issue dates back to the classic work of Coase (1972), who considered a monopolist selling a durable good over time. Coase argued such a monopolist cannot implement a price-skimming strategy when customers rationally anticipate lower future prices and are willing to wait. In fact, with the possibility of continuous price adjustments, he showed that the only SPE price the monopolist can sustain is a uniform price equal to marginal cost. Discount factors play a significant role in the Coase outcome; if the firm is more patient than customers, it may be able to sustain price discrimination (c.f. von der Fehr and Kühn 1995).

Besanko and Winston (1990) show there exist SPE prices that decline over time when both the firm and customers discount utility over time; high-value customers buy early at higher prices whereas low-value customers delay their purchases to obtain lower prices. They show that the SPE price in each period is less than the single-period profit-maximizing price and the seller is worse-off as the number of periods increases. Both results show that the SPE outcome hurts the seller. Other papers that assume dynamic pricing strategies must follow SPE solutions include Cachon and Swinney (2007), Denicolò and Garella (1999), Koenigsberg et al. (2008), Gallien (2006), Elmaghraby et al. (2008), and Su (2007). Aviv and Pazgal (2008) consider two types of pricing strategies including an inventory-contingent markdown policy and preannounced fixed-discount policy. They found that the fixed-discount policy, in general, outperforms contingent pricing under strategic customer purchasing behavior. Similarly, Dasu and Tong (2006) study both contingent pricing and posted pricing schemes, but their focus is on posted pricing. Their extensive numerical study shows that the posted pricing scheme with a small number of price changes is nearly optimal. But the opposite phenomenon is observed by Cachon and Swinney (2007); they find that dynamic pricing outperforms a fixed price path in most cases. One reason in their case is that the firm does not mark down the price in the static pricing setting of their model. To induce strategic customers to purchase in the full price period, the firm has to forgo the benefits of salvaging inventory.

Although SPE behavior is often an appropriate assumption, when a firm has repeated interactions with its customers SPE predictions may be overly pessimistic. Although a firm may have an incentive to deviate from its announced decisions over time, it suffers in the long run because customers will then resort to the SPE and the resulting future losses could well outweigh any short-term gains. Aviv and Pazgal (2008) show numerically that price commitment can increase expected revenues up to 8.32%. With such significant long-term gains at stake, a firm may be able to credibly commit to its decisions as a consequence of the folk theorem of infinitely repeated games (see, for example, Osborne and Rubinstein 1995). To sustain such credible commitment, however, requires an infinite time horizon and sufficiently low discount rates.

At a more basic level, prices may simply be "sticky," and this may prevent ex post price adjustments. Blinder et al. (1998) discuss twelve theories about why prices in many industries are sticky. Among them they note that implicit contracts (both a firm and customers know that the firm will not unilaterally change prices) are an important source of price commitment that is especially relevant with repeat customers. In addition, many sellers commit to prices simply because of advertising constraints or to simplify administration of prices. For example, Broadway theaters sell discounted tickets on the day of performance through TKTS outlets, where their "half-price tickets" policy is central to the concept. Filene's Basement has made a tradition of having automatic markdowns, in which products are marked down based on a preset schedule that begins with 25% off and drops to 75% off after four weeks. Standby airline fares at fixed discounts off full fares are yet another example of such price commitments.

We assume that the selling firm commits to prices and quantities. The assumption is reasonable in many retail settings because of the repeated interaction argument given above; most stores sell products year after year, season after season and are therefore sensitive to future profits and the long-term impact of their pricing policies. We provide formal conditions under which the firm's commitment to prices is credible when the firm sells products repeatedly over an infinite horizon in Online Appendix B (provided in the e-companion).[1] In situations where price and quantity commitment is not a reasonable assumption, our results are, of course, less relevant.

Recently, Yin et al. (2007) and Elmaghraby et al. (2008) also considered the case of a prefixed markdown price path. However, the focus of the former paper is how a firm uses different in-store display formats (display all units or display one unit) to affect customer expectations on the availability. The latter paper focuses on the impact of an option in the full-price period to reserve a product in the sale period if it is not sold yet. Su and Zhang (2007) examine the impact the strategic behavior has on supply chain performance. They show that price and quantity commitment can be achieved using appropriate contractual arrangements as coordinating mechanism.

Our model also does not apply to consumable goods like gasoline, electricity, and groceries, which are used up steadily and replaced periodically, although customers may still behave strategically in purchasing such goods. For example, they may stockpile them during promotions to take advantage of low prices. An extensive stream of research in the marketing science community covers dynamic pricing of consumer nondurables (see Ho et al. 1998, Assunção and Meyer 1993, Bell et al. 2002).

1.1.2. Price or Quantity as a Decision Variable. Our work considers the stocking quantity as the primary tactical decision variable. One motivation for this choice is to understand tactical rationing decisions in the case where firms commit to prices, as in the TKTS and Filene's Basement examples. On a theoretical level, our analysis also provides a complement to traditional inventory theory, which views stocking decisions as primarily driven by holding and fixed costs, overage and underage costs, demand uncertainty, service level constraints, etc. In our model in contrast, the main motivation for stocking decisions is to control the rationing risk faced by customers in order to profitably influence their strategic behavior. In this sense, it provides a behavioral rather than cost-based explanation of stocking decisions.

To our knowledge, almost all papers on dynamic pricing with strategic customers use price as the decision variable. An exception is the work of Dana and Petruzzi (2001), who extend the classical newsvendor model by considering demand to be affected by both price and inventory level (fill rate). In their model, strategic customers face a binary decision—to purchase or not. Because there is a cost to visit the store, their choices depend on both price and the anticipated fill rate. They find that higher fill rates (compared to the traditional newsvendor) are optimal when the firm internalizes the effect of its inventory on demand. Besides the exogenous price case, they explore the case where both price and stocking quantity are decision variables. Su (2007) also considers

[1] An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

optimizing over the initial capacity in his model. He shows the optimal strategy in this case falls into one of three cases, two of which correspond to selling at uniform prices (high or low), and one which corresponds to a single-markdown (high-then-low price) policy. When customers are risk neutral, we find that a similar strategy of selling to all customers at either a single high or low price is optimal. The recent paper by Cachon and Swinney (2007) also allows the firm to choose the initial stocking level and markdown price dynamically. Their model considers a mixture of three types of customers including myopic, strategic, and bargain-hunting customers. With such heterogeneity in behavior, they find that a firm can be better off to adopt a dynamic markdown pricing policy even if it can commit to prices. Ovchinnikov and Milner (2005) investigate how much capacity a firm should allocate for a last minute deal over a series of selling periods to most profitably influence customers' future behavior.

While we focus on capacity decisions, we also address the situation where a firm optimizes over both prices and capacity. In this case, we show that if customers have any degree of risk aversion, then it is always beneficial for the firm to sell at different prices and create some degree of rationing. Thus, we show that with the ability to fine tune price schedules, rationing enables the firm to price discriminate and always emerges as an optimal selling strategy.

1.1.3. Capacity Constraints and Rationing. Whether capacity is exogenously given or endogenously derived, it has a significant influence on pricing strategy. Harris and Raviv (1981) find that the optimal pricing mechanism consists of a uniform price if capacity is not binding, whereas a differential, priority pricing scheme becomes optimal once capacity constraints are binding. In practice, capacity is frequently limited due to budget constraints, finite storage space, precommitments due to long procurement lead time, etc. It may also be the case that capacity is limited deliberately to induce high-value customers to buy at higher prices to avoid rationing risk, as in the case of Zara mentioned above. Indeed, Miguel Diaz Miranda, a Vice President at Zara, explains their strategy as follows:

> Sometimes we make a decision that from an economic point of view might not seem sound, but we know that. For example, we might have an item that was selling very well, but if we think that we are saturating the market with that look we will stop manufacturing it and create unsatisfied demand on purpose. From a strictly economic point of view, that is ridiculous. But the culture we are creating with our customers is: you better get it today because you might not find it tomorrow. (Fraiman et al. 2008, p. 6)

But how should a firm balance the costs and benefits of creating scarcity? This trade-off is a central focus of our work.

Su (2007) also explicitly considers rationing effects and in his model they serve the same role of creating incentives for high-value customers to buy early at high prices. The main modeling difference is that in our paper rationing is implicitly determined by the firm's initial capacity choice, whereas Su introduces an exogenous rationing fraction as an explicit decision variable in his model. Denicolò and Garella (1999) also consider rationing in a model where firms produce and set prices over two periods. They show that rationing can be used to support increased prices over time and improve profits in a durable goods and Coase-type setting. Similarly, Gallego et al. (2004) show that "scraping" unsold products can be strategically optimal when customers expect future sales based on experience. Cachon and Swinney (2007) also find that a firm stocks less with strategic customers than without them in the uncertain demand case. More interestingly, they study the additional value of quick response to mitigate the negative consequences of strategic purchase behavior of customers.

1.1.4. Strategic Interaction Among Customers. A customer's decision may also be affected by the purchase behavior of other customers. This sort of strategic interaction among customers is a key feature of the theory of auction and optimal mechanism design (see Myerson 1981). The important modeling question here is whether strategic interaction among customers should be taken into account when analyzing dynamic pricing. The answer ultimately depends on the market size. If the market consists of a small number of customers whose individual actions significantly affect each other, then modeling the strategic interaction among customers is vital for understanding their behavior. However, in markets consisting of a large number of customers, one customer's behavior has a negligible impact on the outcomes experienced by others. Hence, strategic interactions among customers can be reasonably ignored. It is then a matter of whether one is interested in analyzing a small or large market; our work assumes a large market.

Strategic interaction among customers is considered by Elmaghraby et al. (2008), Gallien (2006), Harris and Raviv (1981), Aviv and Pazgal (2008), Cachon and Swinney (2007), Koenigsberg et al. (2008), Yin et al. (2007), and Dasu and Tong (2006). In contrast, Besanko and Winston (1990), Dana and Petruzzi (2001), Denicolò and Garella (1999), Ovchinnikov and Milner (2005), Levin et al. (2006), and Su (2007), and our work, assume a market consisting of a large customer population and, thus, strategic interaction among customers is ignored. Levin et al. (2006) assume each customer has his individual shopping intensity as a function of prices and each one equally contributes to the overall shopping intensity, which induces a homogenous customer population. They introduce the eagerness of purchase decision as customer's response to prices, and derive the optimal dynamic pricing policy in a stochastic game-theoretical dynamic pricing model.

A large-market assumption simplifies the analysis of customer behavior considerably, because one can then assume each customer reacts to prices and quantities without considering the effect their actions have on other customers. The assumption is also quite reasonable for most mainstream retail categories; as customers, most of us do not worry that our individual shopping decisions will alter the behavior of other "competing shoppers," even though we may account for the aggregate effect of other customers, for example when assessing the likelihood of future shortages. Such behavior is consistent with the large-market assumption.

1.1.5. Risk Preferences. Risk aversion is a reasonable assumption when modeling customer behavior, especially when customers make "large" purchases—purchases whose cost represents a significant portion of a customer's wealth or budget. For example, computers, major household appliances and luxurious fashion items can be regarded as "large" purchases in this sense. Different assumptions of risk preferences can lead to different conclusions about pricing strategies, a fact illustrated by auction theory; a risk-neutral firm prefers a first-price auction to a second-price auction in the face of risk-averse bidders, while it is indifferent to these two formats with risk-neutral buyers.

Our work compares stocking decisions in the presence of risk-neutral and risk-averse customers, respectively. Allowing customers to be risk averse is one distinct feature of our work and the assumption plays a key role in our analysis.

1.1.6. Other Assumptions. Besides the assumptions discussed thus far, the following are also worth noting: First, do customers arrive sequentially or simultaneously? Many papers—including ours—assume that all customers are present at the beginning of the sales season. Aviv and Pazgal (2008), Gallien (2006), Su (2007), Zhou et al. (2006), Yin et al. (2007), and Elmaghraby et al. (2008), in contrast, consider sequential arrivals. Zhou et al. (2006) base their model on that of Gallego and van Ryzin (1994), but allow one customer to strategize over the timing of purchase. Their focus is on the optimal purchasing response of such a strategic customer and its impact on the firm's revenue. Generally, achieving price discrimination is more difficult with simultaneous arrivals because customers have more flexibility in their purchase timing in this case.

Second, is utility constant or discounted over time? In reality, utility is of course discounted. But to simplify analysis, it is acceptable to regard utility as constant if the time horizon is not too long. In Aviv and Pazgal's (2008) work, utility is discounted in the sense that they assume customer valuations strictly decline over time. Gallien (2006) studies an online commerce mechanism design in an infinite time horizon, in which case the use of discounting is natural. In Besanko and Winston's (1990) paper, discount factors play a key role in the SPE; in fact, in their model if utility is not discounted over time, all customers simply purchase in the last period at the lowest price as in the Coase problem. Our basic model does not consider discounting, though we include it as an extension. We ignore discounting partly as a practical approximation, but more importantly to isolate and study the effect that rationing risk alone has on customer behavior.

Last, is aggregate demand deterministic or stochastic? In the setting of uncertain demand, it becomes much harder to model customer strategic behavior. Still, quite a few papers including Aviv and Pazgal (2008), Cachon and Swinney (2007), Ovchinnikov and Milner (2005), Zhou et al. (2006), and Dana and Petruzzi (2001) address dynamic pricing under uncertain aggregate demand. Our work focuses on the case of deterministic aggregate demand, but we address stochastic demand as an extension in Online Appendix A (provided in the e-companion).

1.2. Overview of Our Paper

The remainder of this paper is organized as follows. In §2, we formulate a single seller's stocking decision problem under deterministic aggregate demand. In §3, we characterize the firm's optimal stocking quantity, and examine how the amount of rationing risk is affected by the magnitude of price changes over time and the degree of risk aversion among customers. We also provide conditions under which the firm's commitment to rationing is credible. In §4, we extend the model and analysis to the case where the firm is able to optimize over both prices and capacity, the case where the firm and customers discount profits and utilities over time, and the case of symmetric competing firms in a market where customers assess the aggregate industry availability (rather than firm-level availability) in making their decision to buy early or wait. Conclusions are given in §5. All the proofs are relegated to Online Appendix C (provided in the e-companion).

2. The Model

A monopoly firm preannounces a single-markdown pricing policy over two periods; the unit price in period 1, denoted by p_1, is greater than the unit price in period 2, denoted by p_2. We assume the seller is able to commit to this preannounced price path. We consider a market which consists of a large customer population. The market size, denoted by N, is deterministic. Customers have heterogeneous valuations and unit demand for the good. All customers are present when sales begin and remain in the market until their requests are satisfied or the sales season is over. Customers behave strategically and take both the current and the future prices and availability into consideration when deciding to buy early or late. The firm seeks to maximize profits by choosing its stocking quantity (capacity) at the beginning of the sales season. Inventory is not replenishable once sales start.

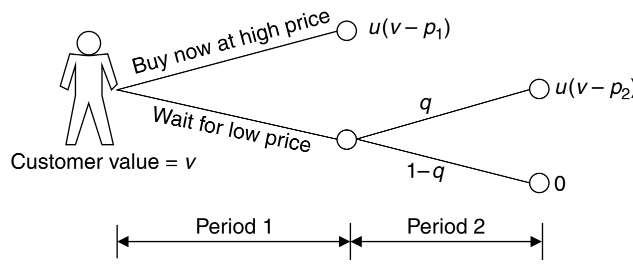
We assume the firm is risk neutral and customers are risk averse. (We study risk-neutral customers as a limiting case.) Customers' valuations are distributed independently and identically with cumulative distribution function $F(v)$, which is common knowledge to both the firm and customers. Moreover, the distribution of customer valuations is constant over time. Customers have identical utility functions, denoted by $u(\cdot)$, which are time invariant, strictly increasing and concave, and twice differentiable; and $u(0) = 0$. We assume there is no disutility when customer demand is unsatisfied; that is, the costs of creating rationing in our model are only lost sales and do not include any loss of goodwill. This is reasonable if customers have good outside alternatives, as in many retail settings. It would not be difficult to add a per unit goodwill cost to our model, but to keep the model parsimonious we choose not to include such costs.

All purchase requests in period 1 are filled, although customers may face a rationing risk in period 2 due to insufficient supply. (We justify these assumptions below.) Let q denote the probability of obtaining a unit in period 2 (the *fill rate*). Random (parallel) rationing is assumed; that is, each customer attempting to purchase in period 2 has an equal chance of obtaining a unit. The fill rate q is determined by the firm's capacity choice, as we show below. We assume customers can correctly anticipate the firm's fill rate; that is, customers have full information and rational expectations.

2.1. The Customer's Decision

As illustrated in Figure 1, a customer weighs the payoff of an immediate purchase at a high price against the expected payoff of a later purchase at a low price and buys one unit in period 1 if and only if $u(v-p_1) \geq qu(v-p_2)$, and $v-p_1 \geq 0$. Intuitively, high-valuation customers are more likely to purchase immediately because they face a larger loss of utility if they are rationed out of the market. The following proposition formalizes this property:

Figure 1 **Two-Period Decision Model**



Proposition 1. $\forall q \in [0,1)$, *there exists the unique* $v(q) \geq p_1$, *such that customers with valuations greater than* $v(q)$ *buy in period* 1 *and those with valuations less than* $v(q)$ *wait to buy in period* 2.

We assume that all customers have the same estimate of fill rates, which is a consequence of the assumption that customers have fully rational expectations. Proposition 1, in fact, characterizes the optimal response of each customer to the firm's capacity choice. At zero fill rate, all customers with valuations greater than p_1 buy at the high price in period 1; thus, $v(0) = p_1$. When the fill rate is one (no rationing risk), no customers purchase in period 1 and we define $v(1) = +\infty$. In the case of finite customer valuations, $v(1) = \bar{U}$, where $\bar{U}$ is the upper bound of valuation. Intuitively, with a higher fill rate, more customers are willing to risk purchasing later at low prices. This is verified by the following proposition:

Proposition 2. *The threshold function* $v(q)$ *is strictly increasing in* q. *Moreover, it is convex on* q *if the associated utility function* $u(\cdot)$ *has the nonnegative third derivative.*[2]

Proposition 2 implies that a lower fill rate (larger rationing risk) induces more customers to buy early at high prices. While a firm would like to induce customers to purchase early at high prices, this increase comes at the expense of lost sales due to rationing in period 2. This trade-off between the benefits of inducing early purchases and the costs of lost sales in period 2 is key to understanding the firm's optimal stocking decisions.

2.2. The Firm's Stocking Decision

Let C be the firm's stocking quantity before sales, and α be the unit procurement cost, $\alpha < p_2$. We assume the firm's cost function is linear; hence, αC is the cost of stocking C units.

[2] Nonnegativity of the third derivative of a utility function is called *prudence* in the economics literature (see Eeckhoudt et al. 1995). For example, the power function $u(x) = x^\gamma$ $(0 < \gamma < 1)$ and the exponential function $u(x) = 1 - e^{-kx}$ $(k > 0)$ are concave and have positive third derivatives.

The fill rate in period 2 is given by the ratio of residual capacity to residual demand in period 2, namely,

$$q = \min\left\{\left(\frac{C - N\bar{F}(v(q))}{N(F(v(q)) - F(p_2))}\right)^+, 1\right\}, \tag{1}$$

where $(\cdot)^+ = \max\{\cdot, 0\}$. Equation (1) shows how the firm is able to influence the fill rate via its capacity choice. The fill rate in turn influences customer behavior through the threshold function $v(q)$, defined (implicitly) by the indifference point:

$$u(v - p_1) = qu(v - p_2). \tag{2}$$

The firm would like to choose its capacity to induce the most profitable demand outcome.

An important fact to note is that, in general, a single capacity choice could induce multiple outcomes. An example is as follows: suppose $u(x) = \sqrt{x}$, $p_1 = 1$, $p_2 = 0.2$, $\alpha = 0$, $C = \frac{3}{4}N$, and customers' valuations are distributed according to $F(v)$, where $F(v) = \frac{1}{4}v^2$, $v \in [0, 2]$. It is easy to verify that both $q = 0$, $v(q) = 1$ and $q = 0.72$, $v(q) = 1.865$ are possible outcomes; that is, both pairs are solutions satisfying (1) and (2). Intuitively, the notion here is that of a "self-fulfilling prophecy"; in the case $q = 0$, $v(q) = 1$, all customers expect a zero fill rate in period 2 ($q = 0$) and those with valuations above one attempt to buy in period 1 as a result ($v(q) = 1$). The resulting demand entirely consumes the capacity in period 1, which indeed produces a fill rate of zero in period 2. On the other hand, if all customers expect a high fill rate in period 2 ($q = 0.72$), many opt to postpone purchasing (those with values below $v(q) = 1.865$). Demand in period 1 is therefore low; hence, there is excess product left over for period 2, which indeed produces a high fill rate of 0.72 in period 2. Just as with multiple equilibria in a game theory model, the predictions of our model are ambiguous in such cases. We will specialize the assumptions below to eliminate this ambiguity, though it is worth keeping in mind that multiple outcomes of this sort could very well be a realistic phenomenon.

We now specify the firm's optimization problem. First, we claim it is optimal for the firm to stock at least enough to meet potential demand at the high price, $N\bar{F}(p_1)$, and never more than required to satisfy the potential demand at the low price, $N\bar{F}(p_2)$. This follows because if the supply is not sufficient to meet demand at the high price, the fill rate in the second period is zero and there will even be rationing at the high price in period 1. Profits in this case are always dominated by choosing $C = N\bar{F}(p_1)$ and serving the entire market at the high price. Conversely, if the firm stocks more than demand at the low price, the fill rate in period 2 will be one, $N\bar{F}(p_2)$ customers will buy in period 2, and there will be positive leftover stock. In this case, profits are always higher by stocking exactly $C = N\bar{F}(p_2)$. Therefore, the optimal capacity choice of the firm always satisfies $C \in [N\bar{F}(p_1), N\bar{F}(p_2)]$, and the fill rate q defined in (1) can therefore be simplified to

$$q = \frac{C - N\bar{F}(v(q))}{N(F(v(q)) - F(p_2))}. \tag{3}$$

We assume that customers' valuations are bounded above by $\bar{U}$. Then according to (2), the market can be segmented only if the fill rate is less than $\bar{q}$, which is defined by $u(\bar{U} - p_1)/u(\bar{U} - p_2)$. Once the fill rate exceeds $\bar{q}$, no customer buys at the high price. As a result, the firm stocks $N\bar{F}(p_2)$ exactly serving the entire low-price market. Therefore, the firm's stocking decision problem is separated to two cases—a segmented market with rationing and a nonsegmented market without rationing. If the market can be segmented, the firm's profit maximization problem can be expressed in terms of C and v as follows:

$$\begin{aligned} \max \quad & N(p_1 - p_2)\bar{F}(v) + (p_2 - \alpha)C \\ \text{s.t.} \quad & u(v - p_1) = \frac{C - N\bar{F}(v)}{N(F(v) - F(p_2))} u(v - p_2), \\ & p_1 \le v \le \bar{U}, \end{aligned} \tag{4}$$

where the first constraint defines the threshold value v using (2) to define v and (3) to define the fill rate.

Let (v^0, C^0) denote the optimal solution to (4), and Π^0 be the associated optimal profit for a segmented market. Denote the profit obtained by serving the entire market at a low price only by Π^{NS}, then $\Pi^{NS} = (p_2 - \alpha)N\bar{F}(p_2)$. So, the firm's optimal stocking quantity corresponds to the one that achieves the maximum of Π^0 and Π^{NS}. We denote the optimal cutoff value by v^*, the optimal fill rate by q^*, and the optimal stocking quantity by C^*.

3. Analysis of the Optimal Stocking Decision

A key question is whether it is optimal to create rationing risk or to simply serve the entire market at one price. If rationing is optimal, what level of rationing risk should be created? And how do these answers depend on the factors such as price differences over time and the level of risk aversion among customers? These are main questions addressed in this section.

To facilitate analysis, we assume in the remainder of this paper, that customers' valuations are uniformly distributed over $[0, \bar{U}]$. We also assume customers have a power utility function $u(x) = x^\gamma$ $(0 < \gamma < 1)$, which is a common form in the economics literature and corresponds to the case where customers have decreasing absolute risk aversion.[3] Lower values of γ correspond to more risk aversion. These assumptions simplify the analysis and also eliminate the multiple-outcome problem mentioned above.

3.1. Optimal Stocking Decision

Using the power utility function $u(x)=x^\gamma$ $(0<\gamma<1)$, the firm's stocking decision for a segmented market given in (4) becomes

$$\begin{aligned}&\max\ \frac{N}{\bar{U}}(p_1-p_2)(\bar{U}-v)+(p_2-\alpha)C\\&\text{s.t.}\ \left(\frac{v-p_1}{v-p_2}\right)^\gamma=\frac{(\bar{U}/N)C-\bar{U}+v}{v-p_2},\quad \bar{U}\ge v\ge p_1.\end{aligned}\tag{5}$$

The profit can be further expressed only in terms of v as follows:

$$\begin{aligned}\max_{\bar{U}\ge v\ge p_1}\Pi(v)=\frac{N}{\bar{U}}\Bigg(&(p_1-\alpha)(\bar{U}-v)\\&+(p_2-\alpha)(v-p_2)\left(\frac{v-p_1}{v-p_2}\right)^\gamma\Bigg).\end{aligned}\tag{6}$$

The first order conditions yield

$$\left(\frac{v-p_1}{v-p_2}\right)^\gamma\left(1+\frac{\gamma(p_1-p_2)}{v-p_1}\right)-\frac{p_1-\alpha}{p_2-\alpha}=0.\tag{7}$$

We then have the following lemma:

Lemma 1. *The profit function $\Pi(v)$ defined in* (6) *is strictly concave in $v\ge p_1$. Furthermore, the maximizer of* (6) *is either the solution to* (7) *denoted by v^0 if $v^0\le\bar{U}$ or $\bar{U}$ otherwise.*

The proof of Lemma 1 establishes that the left-hand side of (7) strictly decreases in $v>p_1$, and has opposite signs at $v\to p_1^+$ and $v\to+\infty$. Hence, there exists a unique root $v^0>p_1$ to (7). Notice also that v^0 is independent of $\bar{U}$. As discussed above, the firm's profit is maximized either at the segmented market optimum or at the low-price-only solution. The following proposition characterizes the optimal solution.

Proposition 3. *Let v^0 be defined as the solution to* (7), *and denote*

$$U_c=\frac{(p_2+\gamma(p_1-\alpha))v^0-p_2(p_1+\gamma(p_2-\alpha))}{v^0-p_1+\gamma(p_1-p_2)}.\tag{8}$$

If $\bar{U}\ge U_c$, the optimal stocking strategy is to induce segmentation by creating rationing risk. The optimal solution in this case is $v^*=v^0$, $q^*=q^0=((v^0-p_1)/(v^0-p_2))^\gamma$, and $C^*=C^0=(N/\bar{U})(\bar{U}-v^0+(v^0-p_2)q^0)$. Otherwise, it is optimal to serve the entire market at the low price; namely, $v^*=\bar{U}$, $q^*=1$, and $C^*=(N/\bar{U})(\bar{U}-p_2)$.

[3] We also numerically tested other distributions of valuations, such as normal and log normal, and other utility functions, including exponential and log. The results are presented in §4.3.

Proposition 3 shows that whether it is optimal to create rationing risk or not depends on the number of high-value customers in the market. When there are a large number of high-value customers ($\bar{U}\ge U_c$), the incremental demand induced in period 1 more than compensates for the lost sales cost of rationing in period 2. If there are relatively few high-value customers ($\bar{U}<U_c$), the opposite is true; the incremental demand induced in period 1 by creating rationing risk does not compensate for the lost sales in period 2. Also, note that, it is never optimal to serve the market only at the high price; the proof of Proposition 3 indicates that the segmented solution (v^0, q^0, C^0) always dominates the high-price-only solution. However, if the firm is able to optimally choose prices for each period, the results are different. Proposition 9 in §4.1 shows that when prices are selected optimally, rationing is always an optimal strategy for any value $\bar{U}$.

One can check U_c defined by (8) decreases in v^0. As shown in the proof of Proposition 3, we know $p_1<v^0<p_1+\gamma(p_2-\alpha)$; therefore, $p_1+\gamma(p_2-\alpha)<U_c<p_1+p_2-\alpha$. This establishes sufficient conditions for creating rationing risk:

Corollary 1. *When $\bar{U}\ge p_1+p_2-\alpha$, creating rationing characterized by (v^0, q^0, C^0) is always optimal. When $\bar{U}\le p_1+\gamma(p_2-\alpha)$, the optimal strategy is to serve the entire market only at the low price.*

Note that $\bar{U}\ge p_1+p_2-\alpha$ implies $(N/\bar{U})(\bar{U}-p_1)\cdot(p_1-\alpha)\ge(N/\bar{U})(\bar{U}-p_2)(p_2-\alpha)$, which says the profit gained at the high-price-only solution is larger than that at the low-price-only solution. Because it is never optimal to serve customers only at the high price in the risk-averse case, this implies when $\bar{U}\ge p_1+p_2-\alpha$ it is always optimal to create rationing risk.

We next consider the limiting case of risk aversion (when γ approaches 1), which corresponds to the case of risk-neutral customers. The result is established in Proposition 4 below.

Proposition 4. *Let v^0 be the solution to* (7), *and $q^0=((v^0-p_1)/(v^0-p_2))^\gamma$. Then $\lim_{\gamma\to1}q^0=0$, and $\lim_{\gamma\to1}v^0=p_1$.*

The above proposition says an extremely high rationing risk is required to induce segmentation as customers become much less risk averse (i.e., $\gamma\to1$). Again, however, whether rationing is optimal or not hinges on the market composition, specifically the number of high-value customers; the above limits only apply to the optimal rationing solution. In summary, for risk-neutral customers, when the market consists of a sufficiently large number of high-value customers ($\bar{U}\ge p_1+p_2-\alpha$), it is optimal for the firm to serve the market only at the high price in period 1; otherwise, the firm serves the entire market at the low price only. Because the high-price-only profit values are the same for both the risk-neutral and risk-averse customers, this shows the firm is strictly better off with risk-averse customers.

3.2. Comparative Statics

In this section, we examine how the optimal fill rate is affected by prices for each period and the degree of risk aversion. We first require the following result:

PROPOSITION 5. *The capacity C increases in the cutoff value v and fill rate q.*

Proposition 5 is intuitively obvious; to produce a higher fill rate (equivalently, a higher cutoff value) requires the firm to stock more. The total sales volume is then increased, but customers' incentive to delay purchasing is increased as well. Again, the firm's decision is to create the "right" rationing risk, balancing the trade-off between benefits of inducing segmentation and costs of lost sales. Proposition 6 and 7 below establish how the optimal amount of rationing risk is influenced by prices for both periods and the degree of risk aversion among customers.

PROPOSITION 6. *The optimal fill rate q^* in the rationing case decreases in the first-period price p_1, while increases in the second-period price p_2.*

This result too is intuitive. As price differences over time decrease (increase the second-period price given that the first-period price is fixed, or reduce the first-period price given that the second-period price is fixed), the opportunity cost of rationing (i.e., the cost of lost sales) increases. On the other hand, a smaller price difference reduces customers' incentive to postpone their purchases. Both effects reduce the benefits of creating rationing in period 2, and drive fill rates up. Note that the optimal fill rate q^* is not necessarily continuous in p_1 and p_2. At some level of p_1 (or p_2), the optimal fill rate jumps to one. This implies inducing segmentation cannot compensate for the lost sales cost at that point.

Another question is how the degree of risk aversion impacts the optimal amount of rationing. Intuitively, the more risk-averse customers are, the more incentive they have to purchase early. Hence, the firm should be able to use a higher fill rate. This is verified in the following proposition:

PROPOSITION 7. *The more risk-averse customers are (smaller γ), the larger the optimal fill rate q^* as long as inducing segmentation via rationing is optimal.*

Proposition 7 implies more rationing risk is needed to induce segmentation as customers become less risk averse. However, at some point, the opportunity cost of rationing may become too great and it then becomes optimal to serve the entire market at the low price. Below we provide the precise condition under which the optimal stocking decision switches from a segmented market to a nonsegmented market, which is established in Proposition 8.

PROPOSITION 8. *Let v^0 be the solution to* (7), *and*

$$q^0 = \left(\frac{v^0 - p_1}{v^0 - p_2}\right)^\gamma, \qquad \hat{q} = \left(\frac{\gamma(p_1 + p_2 - \alpha - \bar{U})}{(\bar{U} - p_2)(1-\gamma)}\right)^\gamma.$$

When $p_1 + \gamma(p_2 - \alpha) < \bar{U} < p_1 + p_2 - \alpha$, q^0 is the optimal fill rate if $q^0 \geq \hat{q}$; otherwise, the optimal fill rate is 1.

3.3. Numerical Examples

In this section, we present several examples to illustrate how the optimal stocking decision is affected by prices for each period (p_1 and p_2) and the degree of risk aversion (γ). We also give examples on the sensitivity of profits to errors in capacity and prices to get a sense of which variable most affects profits.

The left graph in Figure 2 shows that the optimal fill rate q^* decreases in the first-period price p_1 as long as

Figure 2 **Optimal Fill Rates vs. Prices for Each Period**

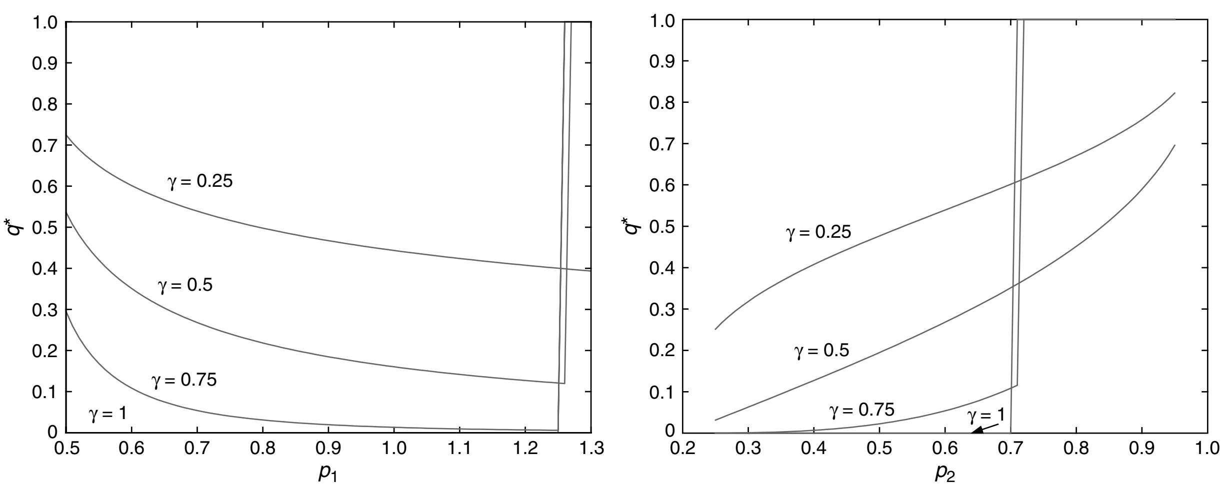
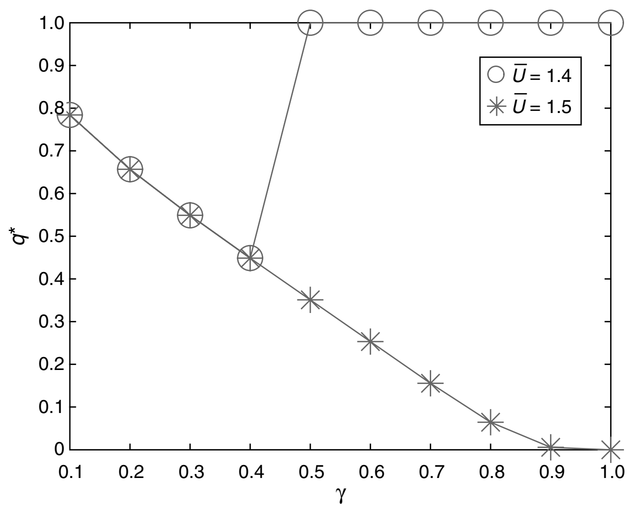


Figure 3 Optimal Fill Rate vs. Degree of Risk Aversion



rationing is optimal; and the right one shows the optimal fill rate q^* increases in the second-period price p_2. In both examples, we set $\bar{U} = 1.5$, $\alpha = 0.2$, and γ takes values over $\{0.25, 0.5, 0.75, 1\}$. The second-period price is set $p_2 = 0.45$ when examining the effect of the first-period price in the left graph, and the first-period price is fixed at $p_1 = 1$ when changing the second-period price in the right graph. Observe that for risk-neutral customers ($\gamma = 1$), the optimal strategy is either a high-price-only solution (when $p_1 \leq 1.25$ in the left graph; $p_2 \leq 0.7$ in the right graph) or a low-price-only solution (when $p_1 > 1.25$ in the left graph and $p_2 > 0.7$ in the right graph) as our theory predicts. Also, we note that the optimal fill rate is not necessarily continuous in prices. At some point, the optimal strategy switches from inducing segmentation to a low-price-only strategy, which, again, reflects the trade-off between benefits of rationing and costs of lost sales.

Figure 3 shows how the optimal fill rate changes with the degree of risk aversion. In this example, we test $p_1 = 1$, $p_2 = 0.7$, $\alpha = 0.2$, $\bar{U} = 1.4$ or 1.5. Observe that when customers are highly risk averse (i.e., γ is small), optimal fill rates are the same for different ranges of valuations. This is because the optimal segmenting point v^0 (or q^0) does not depend on the range of valuations, provided inducing segmentation is an optimal strategy. But, as customers become less risk averse (γ increases), differences emerge. For $\bar{U} = 1.5$, a larger rationing risk continues to be created as customers become less risk averse; in contrast, for $\bar{U} = 1.4$, the firm switches to a low-price-only strategy at some level of risk aversion. In the limiting case ($\gamma \to 1$), the former converges to a fill rate of zero, whereas the latter converges to a fill rate of one, exactly the results for risk-neutral customers.

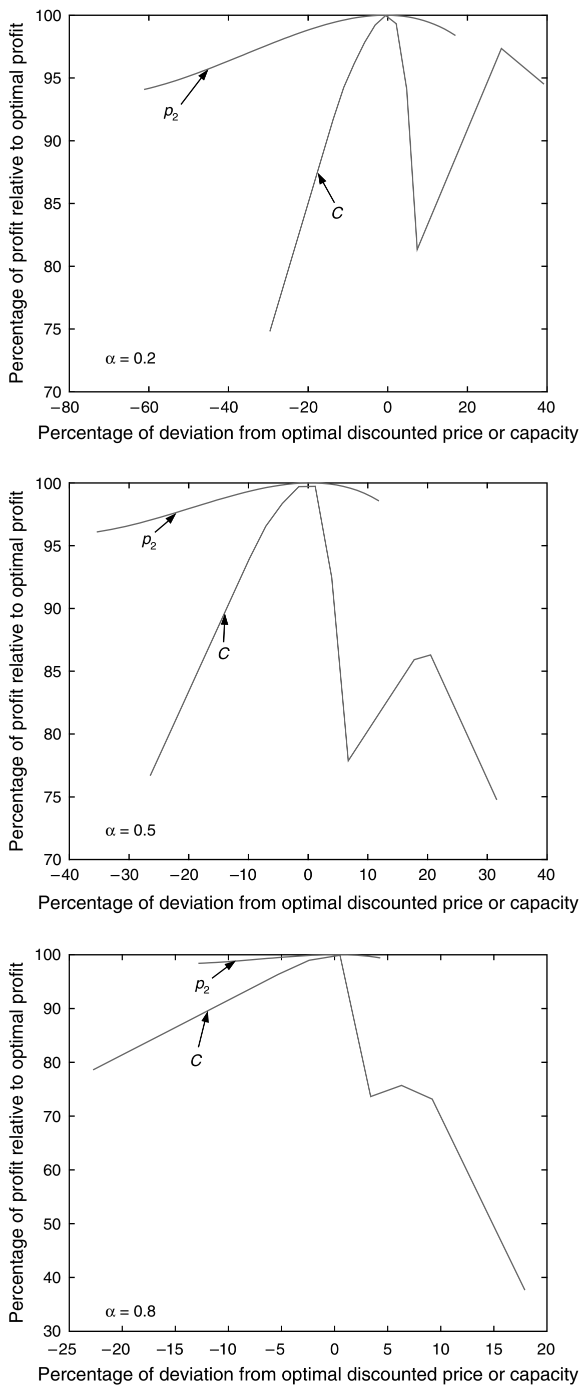
While price is not a variable in our analysis, one might question whether profits are more sensitive to the firm's pricing or capacity decisions. To investigate this question, we consider the situation where the first-period price is given but the firm can choose the discount price in the second period. For example, a manufacturer's suggested retailer price is applied when products are first released to the market, but the firm is able to choose the clearance price on its own. We ran examples on different settings of parameters, and numerically optimized over prices as well as capacities. Given the first-period price, we first found the optimal discount price p_2^* numerically, then calculated the corresponding optimal stocking quantity and profit at p_2^*. Next, we computed relative profit losses under various deviations from p_2^* and from the associated optimal capacity at p_2^*, respectively.

Figure 4 illustrates the results for the case of $N = 1{,}000$, $\gamma = 0.5$, $\bar{U} = 1.5$, $p_1 = 1$, and $\alpha = \{0.2, 0.5, 0.8\}$. In this example, a 40% "error" in the optimal second-period price results in only a 5% relative profit loss (i.e., $1 - [(\text{the optimal profit at } p_2)/(\text{the optimal profit at } p_2^*)] * 100\%$), whereas the same 40% "error" in the optimal capacity results in relative profit losses up to 25%.[4] Profits here are therefore much more sensitive to errors in capacity decisions than errors in pricing decisions. This suggests that inventory policies can be more important than pricing when faced with strategic customers. Why? First, customers make decisions by comparing the current utility, $u(v - p_1)$, with their expected future utility, $qu(v - p_2)$. With a concave utility function, the threshold v may change slowly with changes in p_2; in contrast, a given percentage change in fill rate translates directly into the same percentage change in expected utility in period 2. Second, because fill rate is determined by residual capacity in period 2, a small percentage change in capacity choice can result in a large change in fill rate, thus leading to a large change in the threshold.

Figure 4 also shows that when the price and capacity are chosen poorly, the resulting profit losses can be quite large indeed. Although ideally a firm is able to optimize both prices and capacity, the sensitivity of profit to errors in these decisions illustrates how challenging it is to manage strategic consumers.

[4] The kink in the graph is because the transition from a segmented market to a nonsegmented market induces a jump in capacity. Take the case of $\alpha = 0.2$ for instance, when the deviation from the optimal capacity at p_2^* is less than 7% approximately, the market can be segmented. In this case, the percentage profit relative to the optimal profit increases as a function of the deviation from the optimal capacity, then decreases after the optimal capacity is reached. However, once the percentage deviation from the optimal capacity is larger than 7%, the market cannot be segmented and, hence, the relative profit is increasing in the deviation from the optimal capacity. But after the capacity exceeds the potential demand at the low price, the relative profit decreases again because of the excess supply in period 2.

Figure 4 **Sensitivity of Profits to Errors in Discount Price and Capacity**



4. Extensions to the Basic Problem

We next examine several extensions of our basic problem. The first is to consider the situation where the firm is able to use the optimal prices for each period; that is, the firm optimizes over both prices and capacity and commit to them upfront. We show that a uniform pricing policy is optimal with risk-neutral customers; it is always beneficial for the firm to create rationing if customers are risk averse. Second, we address the case in which the firm and customers discount profits and utilities over time. In general, we show that discounting creates an exogenous incentive to buy early, which adds to the incentive created by rationing risk. Third, we study the robustness of the results by numerically testing other valuation distributions (e.g., normal and log normal) and other utility functions (e.g., exponential and log). We show the main results on the optimal rationing strategy still qualitatively hold for these cases. We also look at the effect of competition, specifically, the case where there are several stores in the market and customers assess the fill rate at the market level rather than the individual store/firm level. We analyze the equilibrium capacity strategies among the firms and show that, in general, competition reduces the industry's ability to profitably use rationing risk to segment customers. Last, we consider the effect of aggregate uncertainty in demand. Though our model of aggregate uncertainty is somewhat artificial, it is amenable to analysis. The general result here is that demand uncertainty reduces a firm's optimal fill rates. The analysis on this case is provided in Online Appendix A.

4.1. Optimizing Over Both Prices and Capacity

When the firm has the ability to choose prices for each period, we solve the optimization problem (6) for a segmented market with p_1 and p_2 as decision variables as well; that is,

$$\max_{\bar{U}\ge v\ge p_1\ge p_2\ge \alpha} \frac{N}{\bar{U}}\bigg((p_1-\alpha)(\bar{U}-v) + (p_2-\alpha)(v-p_2)\bigg(\frac{v-p_1}{v-p_2}\bigg)^{\gamma}\bigg). \quad (9)$$

Note that (6) describes the case in which a market is segmented. If a market is unsegmented, one can easily show that the optimal single price, denoted by p^u, is $(\bar{U}+\alpha)/2$, and the associated profit, denoted by Π^u, is equal to $N(\bar{U}-\alpha)^2/4\bar{U}$.

We first show that with risk-averse customers rationing is always beneficial if the firm is able to use an optimal price schedule. In other words, the value given by (9) is strictly greater than Π^u, the largest profit gained via a single-price market. This is established in Proposition 9 below.

Proposition 9. *The optimal value determined by* (9) *is strictly greater than* Π^u *when* $0<\gamma<1$. *This implies with risk-averse customers, the optimal pricing strategy is a high–low pricing policy* ($p_1>p_2$) *and the optimal stocking decision is to create a fill rate of strictly less than* 1 ($q<1$).

Note that with risk-neutral customers (i.e., $\gamma=1$), a uniform price of $(\bar{U}+\alpha)/2$ is optimal and the associated profit Π^u is strictly less than the profit under risk-averse customers. In essence, Proposition 9 shows that price flexibility and rationing enable the firm to price discriminate and capture surplus from customers who value the product more and prefer to pay to avoid the risk of rationing. It shows that prices can be designed such that the extra revenue from these high-value customers always offsets the cost of rationing. In this sense, the proposition shows rationing is a quite general strategy for extracting surplus and maximizing profits.

Moreover, this result can be generalized to any distribution of valuations under a power utility function. That is, rationing (i.e., to create a fill rate of strictly less than 1) is always an optimal stocking strategy regardless of customer valuation distribution $F(\cdot)$ when $u(x)=x^{\gamma}$ and $0<\gamma<1$.[5]

We ran numerical examples for the endogenous price case similar to those for the basic model. For conciseness we omit the detailed graphs, but the results are qualitatively the same. When prices are endogenous, our numerical examples showed that the optimal fill rate and the optimal first period price decrease in the degree of risk aversion γ; whereas the optimal second period price increases in γ. Recall that when prices are fixed, the optimal fill rate decreases in γ as long as rationing is optimal. We also observed that the relative profit losses under capacity deviations are larger than those under the same percentage of discount price deviations. Therefore, again profits in our numerical tests were more sensitive to errors in capacity than in price.

4.2. Discounted Utility

Even without the risk of rationing, discounting of utilities over time creates an incentive for customers to buy early at high prices. When both discounted utilities and rationing risk are considered, we show that the results are qualitatively the same as those in our basic model without discounting, though characterizing the optimal strategies becomes more complex.

For seasonal, fashionable and durable goods, such as sporting goods, apparel and consumer electronics, it is natural to assume that customers discount the value of such products faster than the firm. We denote the discount factors for the firm and customers by δ_1 and δ_2, respectively, and $\delta_1 \geq \delta_2$. To avoid trivialities, we assume $\delta_1 > \alpha/p_2$ to guarantee a positive margin at the low price. The firm maximizes its total profit, composed of a base profit and an extra profit margin via sales at the high price provided that the market is segmented:

$$\begin{aligned} \max \quad & \frac{N}{\bar{U}}(p_1-\delta_1 p_2)(\bar{U}-v)+(\delta_1 p_2-\alpha)C \\ \text{s.t.} \quad & (v-p_1)^{\gamma}=\delta_2\frac{C-(N/\bar{U})(\bar{U}-v)}{(N/\bar{U})(v-p_2)}(v-p_2)^{\gamma}, \\ & \bar{U}\geq v\geq p_1.^{6} \end{aligned} \tag{10}$$

If the market cannot be segmented, the firm stocks $(N/\bar{U})(\bar{U}-p_2)$ to serve the entire market at one price p_2.

The firm's decision problem for a segmented market given by (10) can be further written in terms of v as follows:

$$\max_{p_1\leq v\leq \bar{U}} \Pi(v) = \frac{N}{\bar{U}}\left\{(p_1-\alpha)(\bar{U}-v)+\frac{1}{\delta_2}(\delta_1 p_2-\alpha) \cdot (v-p_2)\left(\frac{v-p_1}{v-p_2}\right)^{\gamma}\right\}. \tag{11}$$

Following the same approach as in Lemma 1 for the undiscounted case, one can show that the profit function $\Pi(v)$ defined in (11) is strictly concave in $v\geq p_1$. The first order conditions yield:

$$\left(\frac{v-p_1}{v-p_2}\right)^{\gamma}\left(1+\frac{\gamma(p_1-p_2)}{v-p_1}\right)-\frac{\delta_2(p_1-\alpha)}{\delta_1 p_2-\alpha}=0. \tag{12}$$

Applying the same argument as the undiscounted-utility case, one can show there exists a unique solution, denoted by $v^D > 1$ to the Equation (12). Moreover, v^D increases in δ_1 but decreases in δ_2. This result is intuitive. As the firm discounts future profits more (i.e., a smaller δ_1), it has a larger incentive to induce more customers to purchase early; thus, leading to a lower fill rate (note q^D increases in v^D); as the discount factor of customers, δ_2, becomes smaller, future utility becomes less important and more customer have incentive to purchase early; thus, only a smaller rationing risk (i.e., a larger fill rate q^D) is needed.

The precise characterization of the optimal stocking decisions with discounts can be derived using the same approach as in Proposition 3. The results are algebraically complex; specifically,

[5] To see this, replace p_1 by $p_1=p_2 z+v(1-z)$ where $0\leq z\leq 1$, then the firm's profit function can be expressed as $(v-\alpha)N\bar{F}(v)+(p_2-\alpha)N(F(v)-F(p_2))z^{\gamma}-(v-p_2)N\bar{F}(v)z$. When $\gamma=1$, $z^*=0$ or $z^*=1$. This implies no rationing is required (a uniform price is optimal) with risk-neutral customers. When customers are risk averse ($0<\gamma<1$), one can check that $(\tilde{v},\tilde{p}_2,\tilde{z})$ generates a strictly higher profit than a uniform price does, where $\tilde{v}=\arg\max_v (v-\alpha)\bar{F}(v)$, $\tilde{p}_2=\alpha+\epsilon$, $\tilde{z}=(\gamma\epsilon(F(\tilde{v})-F(\tilde{p}_2))/(\tilde{v}-\tilde{p}_2)\bar{F}(\tilde{v}))^{1/(1-\gamma)}$, and ϵ is a small positive number.

[6] We discount utilities here (i.e., a customer with valuation v has a utility in the second period $\delta_2(v-p_2)^{\gamma}$) rather than customers' valuations (i.e., the utility in the second period for this customer is $(\delta_2 v-p_2)^{\gamma}$). We are mainly concerned with the time valuation not the value of the product per se. However, it is quite plausible that the product value declines sharply over time too, in which case having different valuations would be a more natural assumption. We would expect that qualitatively the results would be similar in this case.

Proposition 10. *Let v^D be the solution to the Equation* (12), *and denote*

$$U_c^D = [(\gamma(p_1-p_2)(p_1-\alpha)+p_2(p_1-\delta_1 p_2))v^D - p_1p_2(p_1-\delta_1p_2)-\gamma p_2(p_1-p_2)(\delta_1p_2-\alpha)] \cdot [(p_1-\delta_1p_2)v^D-(p_1-\delta_1p_2)(p_1-\gamma(p_1-p_2))]^{-1}.$$

1. *If $\bar{U}\geq \max\{v^D, U_c^D\}$, inducing segmentation is optimal, namely, $v^*=v^D$, $q^*=(1/\delta_2)((v^*-p_1)/(v^*-p_2))^\gamma$, and $C^*=(N/\bar{U})(\bar{U}-v^*+(v^*-p_2)q^*)$;*
2. *Otherwise, a low-price-only solution is optimal, that is, $v^*=\bar{U}$, $q^*=1$, and $C^*=(N/\bar{U})(\bar{U}-p_2)$.*

The results parallel the undiscounted case. For a large high-value segment, rationing is optimal; for a market with a small number of high-value customers, it is optimal to produce a 100% fill rate.

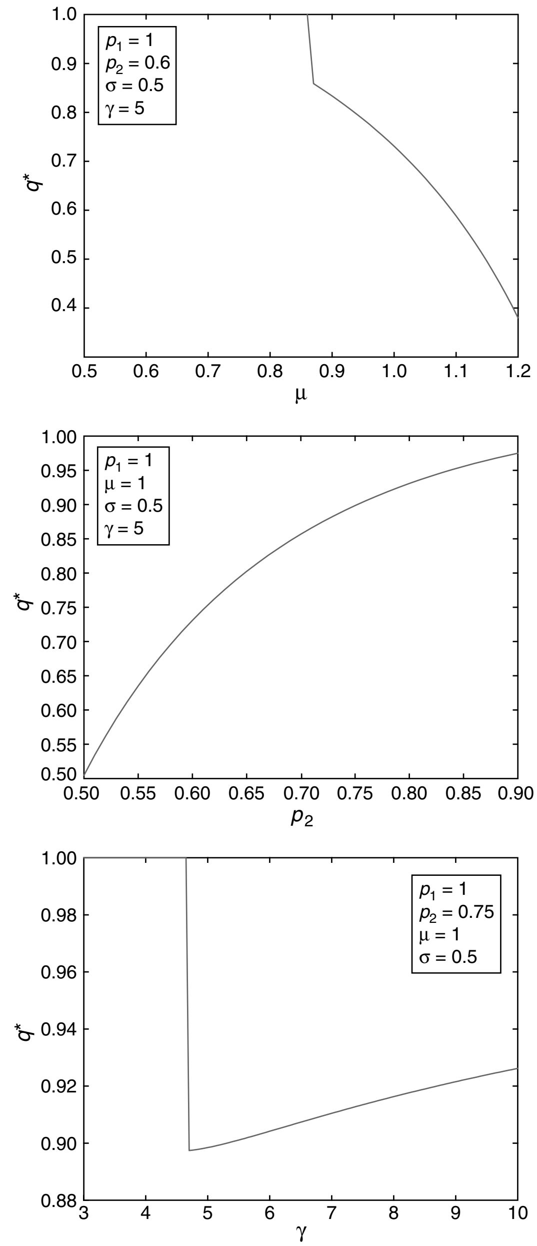
One can easily show that with risk-neutral customers, the optimal stocking decision is either to serve the market at the high price only when $\bar{U}\geq (p_1(p_1-\alpha)-p_2(\delta_1p_2-\alpha))/(p_1-\delta_1p_2)$, or to serve the entire market at the low price when $\bar{U}<(p_1(p_1-\alpha)-p_2(\delta_1p_2-\alpha))/(p_1-\delta_1p_2)$. This result, again, parallels the undiscounted case of risk-neutral customers.

4.3. Other Valuation Distributions and Utility Functions

In our basic model, we assume that customer valuations are uniformly distributed and customers have power utility functions. One may question to what extent the results depend on these particular assumptions. Are the results robust to a general distribution of valuations? Does the form of the utility function matter? Unfortunately, it seems difficult to obtain clean analytical results for more general distributions and utility functions. Yet it is straightforward to analyze these cases numerically. Therefore to assess the robustness of our results, we tested normal and log-normal distributions for customer valuations and exponential and log utility functions. The results indicate that the qualitative behavior of the optimal fill rate as a function of the degree of risk aversion and the second-period price remains the same.

In the interest of space, we illustrate only one case here for the normal distribution of valuation and exponential utility. In this example, customer valuations follow a truncated normal distribution (μ,σ^2) on the support of $[0,\bar{U}]$, and the exponential utility function has the form of $u(x)=1-e^{-\gamma x}$. Note that the exponential utility function $u(x)$ has an Arrow-Pratt measure of absolute risk aversion[7] equal to γ regardless of the value x. The larger the γ, the more risk averse the customers. Figure 5 illustrates how optimal fill rates change with the mean value of customer valuations, the second-period price and the degree of risk aversion, respectively. The results show that the optimal fill rate decreases in the mean value, implying that with a larger high-valuation segment more rationing is optimal as in the case of uniform distribution and power utility. The optimal fill rate increases in the second-period price. Again, this is the same as in the uniform case. The last graph shows that a smaller rationing risk is required as customers become more risk averse; again, this is the same behavior as in the case of uniform valuation and power utility function.

Figure 5 Numerical Examples Under Normal Distribution of Valuations and Exponential Utility Functions



[7] For a twice-differential utility function $u(x)$, the Arrow-Pratt measure of absolute risk aversion at x is defined by $-u''(x)/u'(x)$.

Although these tests of course do not constitute a general proof, we did test all combinations of these distributions and utility functions and found qualitatively identical behavior to the example above. This is good evidence that the basic insights about when rationing is optimal and how the optimal amount of rationing changes as a function of model parameters is robust to our functional assumptions.

4.4. Oligopolistic Competition

When the same product is carried by multiple retailers, it is unlikely that customers base their strategic behavior on the product's availability in any one store. Rather, it is more plausible that they assess availability across the entire market. In doing so, they will consider the aggregate supply and aggregate demand among all stores in the market. Here, we analyze a model of this situation.

Consider an oligopoly market of n firms providing the same product. We assume customers exhibit no preference over the source of supply and with equal probability, a customer buys a product from any firm as long as there is inventory. Capacity choice in the market is a vector, denoted by $C = (C_1, \ldots, C_i, \ldots, C_n)$. We assume all suppliers set the same prices for both periods and have the same unit of procurement cost. This obviously helps keep the model tractable, but is not unreasonable in a competitive retail market, where retailers frequently stock identical products, sell them at the same suggested retail prices, and at nearly identical costs from manufacturers. All the other notations are the same as in a monopoly case. Moreover, two additional assumptions are made to simplify the analysis of stocking decisions under competition: (i) sales in period 1 are equally shared by all firms; and (ii) a buyer will try other firms in period 2 if the firm he initially selects is out of stock until his request is accepted or the market supply is exhausted. Those assumptions are direct consequences of the assumption that customers randomly select suppliers and they have no preference of one specific supplier over another. These assumptions are reasonable in a commodity market.

Using the same argument as in the monopoly case, one can show that the aggregate equilibrium capacity is always greater than the potential demand at the high price, that is $N\bar{F}(p_1)$, and less than the potential demand at the low price equal to $N\bar{F}(p_2)$. Hence, the aggregate fill rate is determined by $q = (\sum_{i=1}^{n} C_i - N\bar{F}(v))/N(F(v) - F(p_2))$. Note that each firm's capacity choice contributes to aggregate fill rate and thus impacts not only its own market share, but also that of all its competitors. This creates a strategic interaction among the stocking decisions of the n firms in the market. Determining optimal stocking quantities under competition therefore becomes a problem of finding equilibria in their capacity choices. Given the perfect symmetry among firms, in what follows we focus only on symmetric equilibria.

Given the capacity choices of other firms', denoted by $C_{-i} = (C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n)$, firm i maximizes its total profit, which, according to the two assumptions made above, consists of an extra margin on an equal share ($1/n$) of first period sales plus a base margin of $p_2 - \alpha$ on each unit, C_i, firm i stocks:

$$\begin{aligned} \max \quad & \Pi_i(C_i, C_{-i}) = \frac{N}{n\bar{U}}(\bar{U} - v)(p_1 - p_2) + (p_2 - \alpha)C_i \\ \text{s.t.} \quad & (v - p_1)^\gamma = \frac{\sum_{i=1}^{n} C_i - (N/\bar{U})(\bar{U} - v)}{(N/\bar{U})(v - p_2)}(v - p_2)^\gamma, \\ & \bar{U} \geq v \geq p_1. \end{aligned} \tag{13}$$

When the market cannot be segmented, the firms provide $(N/\bar{U})(\bar{U} - p_2)$ to serve the entire market at the low price.

Given C_{-i}, the optimization problem for a segmented market (13) can be expressed only in terms of v:

$$\begin{aligned} \max_{\bar{U} \geq v \geq p_1} \Pi_i(v) = {} & \frac{N}{\bar{U}}\left(p_2 - \alpha + \frac{p_1 - p_2}{n}\right)(\bar{U} - v) \\ & + \frac{N}{\bar{U}}(p_2 - \alpha)(v - p_2)\left(\frac{v - p_1}{v - p_2}\right)^\gamma \\ & - (p_2 - \alpha) \sum_{j=1,\, j \neq i}^{n} C_j. \end{aligned}$$

Again, it is easy to show that firm i's profit function, $\Pi_i(v)$, is strictly concave in $v \geq p_1$. The first order conditions yield:

$$\left(1 + \frac{\gamma(p_1 - p_2)}{v - p_1}\right)\left(\frac{v - p_1}{v - p_2}\right)^\gamma - \left(1 + \frac{p_1 - p_2}{n(p_2 - \alpha)}\right) = 0. \tag{14}$$

Using the same argument as for (7), we conclude there exists a unique solution $\bar{v}^0 > p_1$ to (14).

A precise characterization of symmetric Nash equilibria is presented in Proposition 11 below. In a nutshell, as in the monopoly case, the equilibrium involves rationing and segmentation if the market consists of a sufficiently large high-value population, whereas the equilibrium has no rationing or segmentation if the market has a small number of high-value customers. However, both equilibria may be supportable.

Proposition 11. *Let $\bar{v}^0$ be the solution to* (14) *and $\bar{q}^0 = ((\bar{v}^0 - p_1)/(\bar{v}^0 - p_2))^\gamma$. Let*

$$U_c^1 = \frac{n(p_2 - \alpha)(\bar{v}^0 - p_2)(1 - \bar{q}^0)}{p_1 - p_2} + \bar{v}^0,$$

$$U_c^2 = \frac{p_1 - \alpha}{1 - [(n-1)/n]^{1/\gamma}} + p_2.$$

1. *If $\bar{U} \geq U_c^1$, there exists a symmetric segmented Nash equilibrium; namely, $v^* = \bar{v}^0$, $q^* = \bar{q}^0$, and $C_i^* = (N/n\bar{U})(\bar{U} - \bar{v}^0 + (\bar{v}^0 - p_2)q^0)$, $\forall\, i = 1, \ldots, n$.*

2. *If $\bar{U} \leq U_c^1$, or $U_c^1 \leq \bar{U} \leq U_c^2$ and $\bar{q}^0 < 1 - 1/n$, there exists a symmetric low-price-only Nash equilibrium; namely, $v^* = \bar{U}$, $q^* = 1$, and $C_i^* = (N/n\bar{U})(\bar{U} - p_2)$, $\forall\, i = 1, \ldots, n$.*

When customers are risk neutral (i.e., $\gamma = 1$), the equilibrium is either a high-price-only solution or a low-price-only one. In particular, when $\bar{U} \leq n(p_1 - \alpha) + p_2$, a symmetric low-price-only Nash equilibrium exists; that is, no rationing is created, each firm produces $(N/n\bar{U})(\bar{U} - p_2)$ and the entire market is served only at the low price with a 100% fill rate. When $\bar{U} \geq p_1 + n(p_2 - \alpha)$, a symmetric high-price-only Nash equilibrium exists; that is, the market is served only at the high price and each firm stocks $(N/n\bar{U})(\bar{U} - p_1)$. Note in an oligopoly market ($n > 1$), it is always true that $p_2 + n(p_1 - \alpha) > p_1 + n(p_2 - \alpha)$. Hence, multiple equilibria may exist. Specifically, when $\bar{U} > p_2 + n(p_1 - \alpha)$, the symmetric Nash equilibrium is uniquely attained at a high-price-only market; when $\bar{U} < p_1 + n(p_2 - \alpha)$, the unique symmetric Nash equilibrium is a low-price-only strategy; otherwise, both equilibria are attainable outcomes.

Sufficient conditions of uniqueness of symmetric equilibrium are provided in the following corollary:

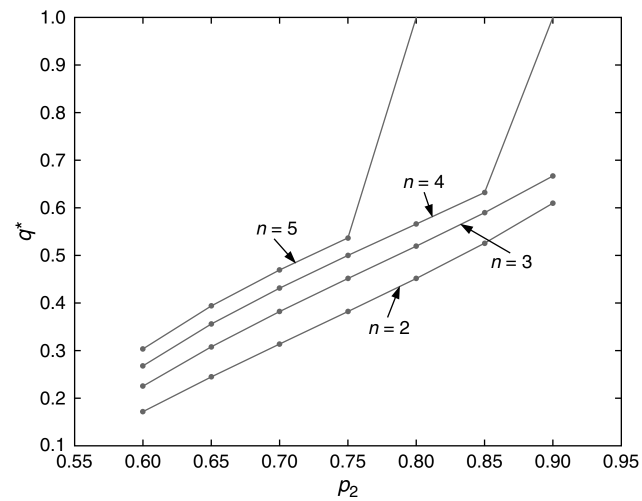
Corollary 2. *If $\bar{U} \geq p_1 + n(p_2 - \alpha)$ and $q^0 \geq 1 - 1/n$, the symmetric Nash equilibrium exists uniquely at a segmented market. If $\bar{U} \leq p_1 + n\gamma(p_2 - \alpha)$, the symmetric Nash equilibrium exists uniquely at a low-price-only solution.*

Intuitively, competition should make a segmentation strategy more difficult to sustain. This is because, with large numbers of competitors, restricting supply has only a negligible impact on the overall market availability but the lost-sales cost of rationing is incurred entirely by firms that are restricting their supply. Indeed, the following corollary shows that there exists a critical number of firms beyond which creating rationing risk is never as sustainable equilibrium. This implies increased competition eventually eliminates the industry's ability to support segmentation via rationing.

Corollary 3. *When the number of firms n becomes sufficiently large, specifically, $n \geq (\bar{U} - p_1)/\gamma(p_2 - \alpha)$, there exists the unique symmetric low-price-only Nash equilibrium.*

One can also compare the outcomes of the oligopoly market with those in the monopoly market. The first difference is that more competition leads to higher aggregate capacity and higher fill rates relative to the monopoly case and these differences increase in the level of competition. However, firms generate lower aggregate profits compared to the monopoly market under more competition, and the difference increases in the level of competition as well.

Figure 6 **Optimal Aggregate Fill Rates Under Competition**



Proposition 12. *The optimal aggregate fill rate, optimal cutoff value, and optimal aggregate capacity in an oligopoly market are larger than those in a monopoly market, whereas the optimal aggregate profit obtained in an oligopoly market is less than that in a monopoly market. Moreover, the optimal aggregate capacity, optimal cutoff value and optimal aggregate fill rate are all increasing in the number of competing firms n; however, the optimal aggregate profit is decreasing in n.*

4.4.1. Numerical Examples. We next illustrate the impact of competition over combinations of the following parameter ranges of $\alpha \in \{0.2, 0.5, 0.8\}$, $\bar{U} \in \{1.25, 1.5, 1.75, 2\}$, and $p_1 = 1$, $\gamma = 0.5$. The examples show that a symmetric segmented Nash equilibrium is more likely to be attained when the marketplace has a small number of competing suppliers. Second, as the number of high-value customers increases (i.e., $\bar{U}$ becomes larger), the equilibrium tends to create rationing risk. For example, at $\alpha = 0.8$ and $\bar{U} = 2$, a segmented Nash equlibrium exists uniquely even at $n = 10$. Figure 6 illustrates how the number of suppliers influences the optimal aggregate fill rate in the case of $\alpha = 0.5$ and $\bar{U} = 2$.

5. Conclusions

Our model shows that rationing can be a profitable strategy to influence the strategic behavior of customers. It also provides a behavioral explanation for stocking and inventory service level decisions that are normally explained in terms of holding and lost-sales cost trade-offs. In our case, the trade-off is between the benefits of inducing customers to purchase early at high prices and the cost of lost sales due to rationing. Our assumptions—a large market of customers who are strategic and risk averse, the ability of the firm to commit to prices and quantities, limited capacity, and a finite selling season—are reasonable as a stylized model for big-ticket, seasonal, and durable goods such as autos, sporting goods, apparel, and consumer electronics. (Albeit our oligopoly model assumes commodity products, which only fit some of these categories.)

Under our assumptions, rationing is not profitable when customers are risk neutral. But even when customers are risk averse, rationing may not be optimal if the number of high-valuation customers is too small. In general, a large high-value customer segment, high levels of risk aversion and large differences in price over time all tend to favor rationing as an optimal strategy. Numerical examples suggest that capacity decisions can be even more important than price in terms of influencing strategic customer behavior. When the firm has the ability to choose prices, however, rationing is always an optimal strategy. Last, our oligopoly analysis shows that competition makes it more difficult to support segmentation using rationing, and when the number of competitors is sufficiently large, a low-price-only equilibrium is the only sustainable outcome. Thus, rationing is more likely to be used in cases where a firm has some reasonable degree of market power.

As for future work, we have a forthcoming paper investigating the case where customers do not perfectly anticipate fill rates, but rather update their estimates over time based on experience. The issue here is understanding how the firm should respond to most profitably influence these expectations over time, whether the market converges to an equilibrium and, if so, how the equilibrium is related to the rational-expectations outcome derived here.

6. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

Acknowledgments

The authors thank two anonymous referees and an associate editor for many constructive suggestions on an earlier draft of this work. In particular, the analysis in §4.1 was suggested by one referee.

References

Assunção, J., R. J. Meyer. 1993. The rational effect of price promotions on sales and consumption. *Management Sci.* **39** 517–535.

Aviv, Y., A. Pazgal. 2008. Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing Service Oper. Management*, ePub ahead of print January 4, http://msom.journal.informs.org/cgi/content/abstract/msom.1070.0183v1.

Bell, D. R., G. Iyer, V. Padmanabhan. 2002. Price competition under stockpiling and flexible consumption. *J. Marketing Res.* **39** 292–303.

Besanko, D., W. L. Winston. 1990. Optimal price skimming by a monopolist facing rational consumers. *Management Sci.* **36** 555–567.

Blinder, A., E. Canetti, D. Lebow, J. Rudd. 1998. *Asking About Prices: A New Approach to Understanding Price Stickiness*. Russell Sage Foundation, New York.

Bulow, J. I. 1982. Durable-goods monopolists. *J. Political Econom.* **90** 314–332.

Cachon, G., R. Swinney. 2007. Purchasing, pricing and quick response in the presence of strategic consumers. Working paper, The Wharton School, University of Pennsylvania, Philadelphia.

Chen, X., D. Simchi-Levi. 2003. Coordinating inventory control and pricing strategies with random demand and fixed ordering costs: The finite horizon case. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge.

Coase, R. H. 1972. Durability and monopoly. *J. Law Econom.* **15** 143–149.

Dana, J. D., N. C. Petruzzi. 2001. Note: The newsvendor model with endogenous demand. *Management Sci.* **47** 1488–1497.

Dasu, S., C. Tong. 2006. Dynamic pricing when consumers are strategic: Analysis of a posted pricing scheme. Working paper, Marshall School of Business, University of Southern California, Los Angeles.

Denicolò, V., P. G. Garella. 1999. Rationing in a durable goods monopoly. *Rand J. Econom.* **30** 44–55.

Eeckhoudt, L., C. Gollier, T. Schneider. 1995. Risk-aversion, prudence and temperance: A unified approach. *Econom. Lett.* **48** 331–336.

Elmaghraby, W., A. Gülcü, P. Keskinocak. 2008. Optimal markdown mechanisms in the presence of rational customers with multi-unit demands. *Manufacturing Service Oper. Management* Forthcoming.

Federgruen, A., A. Heching. 1999. Combined pricing and inventory control under uncertainty. *Oper. Res.* **47** 454–475.

Feng, Y., G. Gallego. 1995. Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management Sci.* **41** 1371–1391.

Ferdows, K., M. A. Lewis, J. A. D. Machuca. 2005. Zara's secret for fast fashion. Retrieved February 21, http://hbswk.hbs.edu/tools.

Fraiman, N., M. Singh, L. Arrington, C. Paris. 2008. ZARA. Columbia Business School Caseworks, Columbia University, New York.

Gallego, G., G. J. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* **40** 999–1020.

Gallego, G., R. Phillips, O. Sahin. 2004. Strategic management of distressed inventory. Working paper, Department of Industrial Engineering and Operations Research, Columbia University, New York.

Gallien, J. 2006. Dynamic mechanism design for online commerce. *Oper. Res.* **54** 291–310.

Harris, M., A. Raviv. 1981. A theory of monopoly pricing schemes with demand uncertainity. *Amer. Econom. Rev.* **71** 347–365.

Ho, T. H., C. S. Tang, D. R. Bell. 1998. Rational shopping behavior and the option value of variable pricing. *Management Sci.* **44** 145–160.

Koenigsberg, O., E. Muller, N. J. Vilcassim. 2008. Easyjet pricing strategy: Should low-fare airlines offer last-minute deals? *Quant. Marketing Econom.* Forthcoming.

Levin, Y., J. McGill, M. Nediak. 2006. Optimal dynamic pricing of perishable items by a monopolist facing strategic consumers. Working paper, Queen's School of Business, Queen's University, Kingston, Canada.

Myerson, R. 1981. Optimal auction design. *Math. Oper. Res.* **6** 58–73.

Osborne, M. J., A. Rubinstein. 1995. *A Course in Game Theory*. MIT Press, Cambridge, MA.

Ovchinnikov, A., J. M. Milner. 2005. Strategic response to wait-or-buy: Revenue management through last minute deals in the presence of customer learning. Working paper, Joseph L. Rotman School of Management, University of Toronto, Toronto.

Rozhon, T. 2005. Before christmas, Wal-Mart was stirring. Retrieved January 5, http://www.nytimes.com/.

Su, X. 2007. Inter-temporal pricing with strategic customer behavior. *Management Sci.* **53** 726–741.

Su, X., F. Zhang. 2007. Strategic customer behavior, commitment, and supply chain performance. Working paper, Haas School of Business, University of California, Berkeley.

Talluri, K. T., G. J. van Ryzin. 2004. *The Theory and Practice of Revenue Management*. Kluwer, Boston.

von der Fehr, N.-H. M., K.-U. Kühn. 1995. Coase versus Pacman: Who eats whom in the durable-goods monopoly? *J. Political Econom.* **103** 785–812.

Yin, R., Y. Aviv, A. Pazgal, C. S. Tang. 2007. Optimal markdown pricing: Implications of inventory display formats in the presence of strategic customers. Working paper, W. P. Carey School of Business, Arizona State University, Tempe.

Zhou, Y., M. Fan, M. Cho. 2006. Strategic consumer response to dynamic pricing of perishable products. Working paper, University of Washington Business School, Seattle.